

Smoothness-constrained model for nonparametric item response theory

Toshiki Sato · Yuichi Takano

Received: date / Accepted: date

Abstract This paper is concerned with the nonparametric item response theory (NIRT) for estimating item characteristic curves (ICCs) and latent abilities of examinees on educational and psychological tests. In contrast to parametric models, NIRT models can estimate various forms of ICCs under mild shape restrictions, such as the constraints of monotone homogeneity and double monotonicity. However, NIRT models frequently suffer from estimation instability because of the great flexibility of nonparametric ICCs, especially when there is only a small amount of item-response data. To improve the estimation accuracy, we propose a novel NIRT model constrained by monotone homogeneity and smoothness based on ordered latent classes. Our smoothness constraints avoid overfitting of nonparametric ICCs by keeping them close to logistic curves. We also implement a tailored expectation-maximization algorithm to calibrate our smoothness-constrained NIRT model efficiently. We conducted computational experiments to assess the effectiveness of our smoothness-constrained model in comparison with the common two-parameter logistic model and the monotone-homogeneity model. The computational results demonstrate that our model obtained more accurate estimation results than did the two-parameter logistic model when the latent abilities of examinees for some test items followed bimodal distributions. Moreover, our model outperformed the monotone-homogeneity model because of the effect of the smoothness constraints.

Keywords Item response theory · Nonparametric estimation · Smoothness constraint · Optimization · EM algorithm · Latent class

Toshiki Sato
Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1
Tennodai, Tsukuba-shi, Ibaraki 305-8573, Japan

Yuichi Takano
School of Network and Information, Senshu University, 2-1-1 Higashimita, Tama-ku,
Kawasaki-shi, Kanagawa 214-8580, Japan

1 Introduction

Item response theory (IRT) is a family of statistical measurement methods for educational and psychological tests. In IRT models, the characteristics of each test item are examined based on the item characteristic curve (ICC), which expresses the probability of a correct answer as a function of the latent abilities of the examinees. Indeed, many testing companies use IRT models for the design, analysis, and scoring of tests.

This paper is focused on nonparametric item response theory (NIRT) models (Junker and Sijtsma 2001; Sijtsma and Molenaar 2002; Stout 1987; van der Linden and Hambleton 2013). In contrast to parametric item response theory (PIRT) models in which the ICCs are defined by parametric functions (e.g., logistic curves or normal ogives), NIRT models are capable of estimating various forms of ICCs under mild shape restrictions, such as the monotone-homogeneity constraint (Meredith 1965; Mokken 1971) and the double monotonicity constraint (Mokken 1971; Mokken and Lewis 1982). It has been demonstrated that PIRT models do not always fit the data well (Douglas and Cohen 2001; Johnson 2007; Ramsay 1991), under which circumstances NIRT models have a clear advantage. They are also useful for evaluating the data quality (Meijer et al. 2015) and the goodness of fit of PIRT models (Douglas and Cohen 2001; Liang and Wells 2015; Liang et al. 2014).

The existing methods for estimating nonparametric ICCs include regression splines (Johnson 2007; Ramsay 1988; Ramsay and Abrahamowicz 1989; Ramsay and Winsberg 1991; Rossi et al. 2002), kernel smoothing (Douglas 1997; Guo and Sinharay 2011; Luzardo and Rodríguez 2015; Ramsay 1991), isotonic regression (Lee 2007), the finite mixture model (Mori and Kano 2013), and monotonic polynomial regression (Falk and Cai 2016; Liang and Browne 2015). If the latent abilities of the examinees are represented as ordered latent classes, NIRT models are categorized as ordered latent class models (Croon 1990, 1991; Ligtoet and Vermunt 2012; van Onna 2002; Vermunt 2001). The expectation-maximization (EM) algorithm (Croon 1990, 1991; Rossi et al. 2002; Vermunt 2001) and Markov chain Monte Carlo (MCMC) method (Johnson 2007; Karabatsos and Sheu 2004; Levy 2009; Ligtoet and Vermunt 2012; Monroe and Cai 2014; van Onna 2002) have been employed to estimate both the nonparametric ICCs and the latent abilities of examinees. However, NIRT models frequently suffer from estimation instability because of the great flexibility of nonparametric ICCs, especially if there is only a small amount of item-response data (Molenaar 2001).

Various shape restrictions have been proposed to prevent overfitting with nonparametric regression models (Ramsay and Silverman 2005; Simonoff 2012). To improve the estimation accuracy of NIRT models, we make effective use of the smoothness constraints on the nonparametric ICCs. More specifically, we propose an NIRT model with monotone homogeneity and smoothness constraints based on the ordered latent classes. Our smoothness constraints keep each nonparametric ICC close to a logistic curve, and thus offer advantages to both PIRT and NIRT models, namely stability and flexibility. To the best

of our knowledge, no existing study has incorporated such smoothness constraints into a monotone-homogeneity NIRT model. In addition, we implement a tailored EM algorithm to calibrate our smoothness-constrained NIRT model efficiently.

We conducted computational experiments to assess the effectiveness of our smoothness-constrained model in comparison with the common two-parameter logistic model and the monotone-homogeneity model. The computational results demonstrate that our model delivered the best estimation performance in many cases. In other words, the smoothness constraints were very effective in enhancing the estimation accuracy of the NIRT models.

The remainder of this paper is organized as follows. In Sect. 2, we present the monotone-homogeneity model for estimating the monotonically increasing ICCs and the ability classes of examinees. In Sect. 3, we formulate our smoothness-constrained model and EM algorithm. In Sect. 4, we report the computational results, and in Sect. 5 we conclude the paper with a brief summary of our work.

2 Monotone-homogeneity model

In this section, we pose our monotone-homogeneity model by following Takano et al. (2016). Let us denote by I a set of examinees and by J a set of dichotomously scored question items on a test. The test results are given as the binary item-response data

$$\mathbf{U} := (u_{ij})_{(i,j) \in I \times J} \in \{0, 1\}^{|I| \times |J|},$$

where $u_{ij} = 1$ if the i th examinee provided a correct answer to the j th item, or $u_{ij} = 0$ otherwise. We make the following assumptions throughout this paper:

Unidimensionality: the latent abilities of all examinees are evaluated unidimensionally.

Local Independence: item responses are conditionally independent of each other given an individual latent ability.

Additionally, the latent abilities of examinees are represented as ordered latent classes denoted by T .

The nonparametric ICCs of test items are defined by the decision variable

$$\mathbf{X} := (x_{jt})_{(j,t) \in J \times T},$$

where x_{jt} is the probability of the j th item being answered correctly by examinees in the t th ability class. These nonparametric ICCs are usually estimated subject to monotone-homogeneity constraints (Meredith 1965; Mokken 1971), which require that the probability of a correct answer increases monotonically with ability class:

$$x_{jt} \leq x_{j,t+1} \quad ((j, t) \in J \times T), \quad (1)$$

$$0 \leq x_{jt} \leq 1 \quad ((j, t) \in J \times T). \quad (2)$$

The ability classes of examinees are represented by the decision variable

$$\mathbf{Y} := (y_{it})_{(i,t) \in I \times T},$$

where $y_{it} = 1$ if the i th examinee possesses the latent ability of the t th class, or $y_{it} = 0$ otherwise. The following constraints guarantee that only one ability class is assigned to each examinee:

$$\sum_{t \in T} y_{it} = 1 \quad (i \in I), \quad (3)$$

$$y_{it} \in \{0, 1\} \quad ((i, t) \in I \times T). \quad (4)$$

Given $\mathbf{x}_{j\cdot} := (x_{jt})_{t \in T}$ and $\mathbf{y}_{i\cdot} := (y_{it})_{t \in T}$, the probability of receiving response $u_{ij} \in \{0, 1\}$ is expressed as

$$\Pr(u_{ij} \mid \mathbf{x}_{j\cdot}, \mathbf{y}_{i\cdot}) := \prod_{t \in T} ((x_{jt})^{u_{ij}} (1 - x_{jt})^{1-u_{ij}})^{y_{it}}.$$

From the assumption of local independence, the probability of the i th examinee giving the response $\mathbf{u}_i := (u_{ij})_{j \in J}$ is expressed as

$$\Pr(\mathbf{u}_i \mid \mathbf{X}, \mathbf{y}_{i\cdot}) := \prod_{j \in J} \Pr(u_{ij} \mid \mathbf{x}_{j\cdot}, \mathbf{y}_{i\cdot}).$$

Because the responses of different examinees are independent, the probability of receiving item response \mathbf{U} from all the examinees is given by

$$\begin{aligned} \Pr(\mathbf{U} \mid \mathbf{X}, \mathbf{Y}) &:= \prod_{i \in I} \Pr(\mathbf{u}_i \mid \mathbf{X}, \mathbf{y}_{i\cdot}) \\ &= \prod_{(i,j,t) \in I \times J \times T} ((x_{jt})^{u_{ij}} (1 - x_{jt})^{1-u_{ij}})^{y_{it}}. \end{aligned}$$

By treating \mathbf{X} and \mathbf{Y} as decision variables, the log-likelihood function is defined as follows:

$$\begin{aligned} \ell(\mathbf{X}, \mathbf{Y} \mid \mathbf{U}) &:= \log \Pr(\mathbf{U} \mid \mathbf{X}, \mathbf{Y}) \\ &= \sum_{(i,j,t) \in I \times J \times T} y_{it} (u_{ij} \log x_{jt} + (1 - u_{ij}) \log(1 - x_{jt})). \end{aligned} \quad (5)$$

Consequently, the monotone-homogeneity model estimates \mathbf{X} and \mathbf{Y} so that the log-likelihood function (5) is maximized subject to constraints (1)–(4):

$$\begin{aligned} &\underset{\mathbf{X}, \mathbf{Y}}{\text{maximize}} && \sum_{(i,j,t) \in I \times J \times T} y_{it} (u_{ij} \log x_{jt} + (1 - u_{ij}) \log(1 - x_{jt})) \end{aligned} \quad (6)$$

$$\text{subject to } x_{jt} \leq x_{j,t+1} \quad ((j, t) \in J \times T), \quad (7)$$

$$0 \leq x_{jt} \leq 1 \quad ((j, t) \in J \times T), \quad (8)$$

$$\sum_{t \in T} y_{it} = 1 \quad (i \in I), \quad (9)$$

$$y_{it} \in \{0, 1\} \quad ((i, t) \in I \times T). \quad (10)$$

3 Smoothness-constrained model

In this section, we firstly formulate our smoothness-constrained model and then describe an EM algorithm for model estimation.

3.1 Smoothness constraints

To express our smoothness constraints, we use the logistic function

$$\lambda(w) := \frac{1}{1 + \exp(-w)} \quad (11)$$

and the additional decision variable

$$\mathbf{W} := (w_{jt})_{(j,t) \in J \times T}.$$

The ICCs are then defined as $x_{jt} = \lambda(w_{jt})$; that is, $\lambda(w_{jt})$ denotes the probability of the j th item being answered correctly by examinees in the t th ability class. Because the logistic function increases monotonically from zero to one, the monotone-homogeneity constraints on $\lambda(w_{jt})$ are written as

$$w_{jt} \leq w_{j,t+1} \quad ((j, t) \in J \times T). \quad (12)$$

The smoothness constraints on the nonparametric ICCs are posed as follows:

$$\sum_{t \in T} |w_{j,t+2} - 2w_{j,t+1} + w_{jt}| \leq \gamma \quad (j \in J), \quad (13)$$

where $\gamma \geq 0$ is a user-defined parameter. If γ is sufficiently large, constraints (13) are invalidated. Conversely, if $\gamma = 0$, constraints (13) are equivalent to

$$w_{j,t+2} - w_{j,t+1} = w_{j,t+1} - w_{jt} \quad ((j, t) \in J \times T), \quad (14)$$

which imply that for each $j \in J$, $\{w_{jt} \mid t \in T\}$ is a set of equally spaced points.

Fig. 1 illustrates three examples of smoothness-constrained ICCs with $\gamma = 0$, where the figures on the left-hand side are graphs of the logistic function, and those on the right-hand side are the corresponding ICCs. It is clear that an S-shaped ICC can be created as shown in Fig. 1(a). In addition, although $w_{j1}, w_{j2}, \dots, w_{j5}$ must be equally spaced points, the difficulty and discrimination of a test item can be adjusted. For instance, Fig. 1(b) and (c) correspond to difficult and undiscriminating items, respectively. These examples demonstrate that the smoothness constraints (13) keep each ICC close to a logistic curve with two parameters, namely difficulty and discrimination. Therefore, our smoothness constraints provide benefits to both PIRT and NIRT models in the sense that the shapes of nonparametric ICCs are restricted by means of parametric functions.

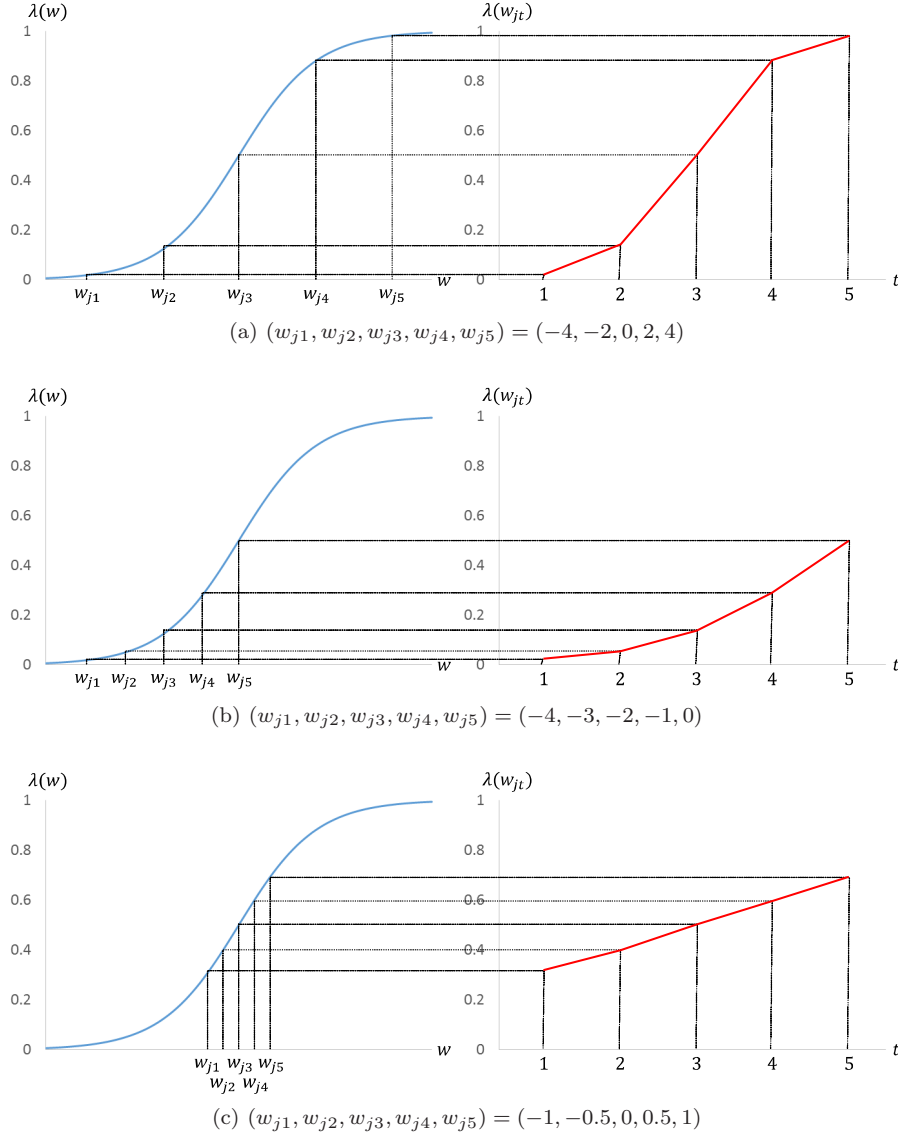


Fig. 1 Smoothness-constrained ICCs with $\gamma = 0$

3.2 Formulation

We begin by substituting $x_{jt} = \lambda(w_{jt})$ into the log-likelihood function (5) as follows:

$$\sum_{(i,j,t) \in I \times J \times T} y_{it} (u_{ij} \log \lambda(w_{jt}) + (1 - u_{ij}) \log(1 - \lambda(w_{jt}))). \quad (15)$$

It then follows from Eq. (11) that

$$\begin{aligned}
& u_{ij} \log \lambda(w_{jt}) + (1 - u_{ij}) \log(1 - \lambda(w_{jt})) \\
&= u_{ij} \log \lambda(w_{jt}) + (1 - u_{ij}) \log \lambda(-w_{jt}) \\
&= \log \lambda((2u_{ij} - 1)w_{jt}) \quad (\because u_{ij} \in \{0, 1\}) \\
&= -\log(1 + \exp((1 - 2u_{ij})w_{jt})).
\end{aligned} \tag{16}$$

Therefore, maximizing the log-likelihood function (15) is equivalent to minimizing

$$\sum_{(i,j,t) \in I \times J \times T} y_{it} \log(1 + \exp((1 - 2u_{ij})w_{jt})). \tag{17}$$

We show next that the smoothness constraints (13) can be converted into a set of linear constraints.

Proposition 1 *The smoothness constraints (13) hold if and only if there exist $\mathbf{S} := (s_{jt})_{(j,t) \in J \times T}$ and $\mathbf{V} := (v_{jt})_{(j,t) \in J \times T}$ such that*

$$\sum_{t \in T} (s_{jt} + v_{jt}) \leq \gamma \quad (j \in J), \tag{18}$$

$$s_{jt} - v_{jt} = w_{j,t+2} - 2w_{j,t+1} + w_{jt} \quad ((j, t) \in J \times T), \tag{19}$$

$$s_{jt} \geq 0, \quad v_{jt} \geq 0 \quad ((j, t) \in J \times T). \tag{20}$$

Proof Firstly, we suppose that the smoothness constraints (13) are satisfied by $\bar{\mathbf{W}} := (\bar{w}_{jt})_{(j,t) \in J \times T}$. We then set $\bar{\mathbf{S}} := (\bar{s}_{jt})_{(j,t) \in J \times T}$ and $\bar{\mathbf{V}} := (\bar{v}_{jt})_{(j,t) \in J \times T}$ as

$$\begin{cases} \bar{s}_{jt} := \bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}, & \bar{v}_{jt} := 0 & \text{if } \bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt} \geq 0, \\ \bar{s}_{jt} := 0, & \bar{v}_{jt} := -(\bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}) & \text{otherwise,} \end{cases}$$

for each $(j, t) \in J \times T$. It then follows that constraints (18)–(20) are satisfied by $\bar{\mathbf{S}}$, $\bar{\mathbf{V}}$, and $\bar{\mathbf{W}}$ because

$$\sum_{t \in T} (\bar{s}_{jt} + \bar{v}_{jt}) = \sum_{t \in T} |\bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}| \leq \gamma \quad (j \in J).$$

Conversely, we suppose that constraints (18)–(20) are satisfied by $\bar{\mathbf{S}}$, $\bar{\mathbf{V}}$, and $\bar{\mathbf{W}}$. The smoothness constraints (13) are then satisfied as follows:

$$\sum_{t \in T} |\bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}| = \sum_{t \in T} |\bar{s}_{jt} - \bar{v}_{jt}| \leq \sum_{t \in T} (\bar{s}_{jt} + \bar{v}_{jt}) \leq \gamma \quad (j \in J).$$

□

Consequently, our smoothness-constrained model minimizes Eq. (17) subject to constraints (3)–(4), (12), and (18)–(20):

$$\underset{\mathbf{S}, \mathbf{V}, \mathbf{W}, \mathbf{Y}}{\text{minimize}} \quad \sum_{(i,j,t) \in I \times J \times T} y_{it} \log(1 + \exp((1 - 2u_{ij})w_{jt})) \quad (21)$$

$$\text{subject to} \quad w_{jt} \leq w_{j,t+1} \quad ((j, t) \in J \times T), \quad (22)$$

$$\sum_{t \in T} (s_{jt} + v_{jt}) \leq \gamma \quad (j \in J), \quad (23)$$

$$s_{jt} - v_{jt} = w_{j,t+2} - 2w_{j,t+1} + w_{jt} \quad ((j, t) \in J \times T), \quad (24)$$

$$s_{jt} \geq 0, \quad v_{jt} \geq 0 \quad ((j, t) \in J \times T), \quad (25)$$

$$\sum_{t \in T} y_{it} = 1 \quad (i \in I), \quad (26)$$

$$y_{it} \in \{0, 1\} \quad ((i, t) \in I \times T). \quad (27)$$

3.3 EM algorithm

We employ an EM algorithm to find a good-quality solution to the smoothness-constrained model (21)–(27) efficiently. To this end, we introduce the decision variable $\boldsymbol{\pi} := (\pi_t)_{t \in T}$ for the sizes of ability classes, such that

$$\sum_{t \in T} \pi_t = 1, \quad \pi_t > 0 \quad (t \in T). \quad (28)$$

The conditional probability of receiving response \mathbf{u}_i from the i th examinee, given that s/he belongs to the t th ability class (i.e., $y_{it} = 1$), is expressed as

$$f(\mathbf{u}_i \mid \mathbf{w}_{\cdot t}) := \prod_{j \in J} \lambda(w_{jt})^{u_{ij}} (1 - \lambda(w_{jt}))^{1-u_{ij}}.$$

Accordingly, the marginal likelihood is calculated by a weighted sum of the form

$$\sum_{t \in T} \pi_t f(\mathbf{u}_i \mid \mathbf{w}_{\cdot t}).$$

The posterior probability of the i th examinee belonging to the t th ability class is given by Bayes' rule as follows:

$$\frac{\pi_t f(\mathbf{u}_i \mid \mathbf{w}_{\cdot t})}{\sum_{\tau \in T} \pi_\tau f(\mathbf{u}_i \mid \mathbf{w}_{\cdot \tau})}. \quad (29)$$

Moreover, the complete-data log-likelihood function based on \mathbf{Y} is formulated as follows:

$$\begin{aligned} & \log \left(\prod_{(i,t) \in I \times T} (\pi_t f(\mathbf{u}_{i\cdot} \mid \mathbf{w}_{\cdot t}))^{y_{it}} \right) \\ &= \underbrace{\sum_{(i,t) \in I \times T} y_{it} \log \pi_t}_{(30a)} + \underbrace{\sum_{(i,t) \in I \times T} y_{it} \log f(\mathbf{u}_{i\cdot} \mid \mathbf{w}_{\cdot t})}_{(30b)}. \end{aligned} \quad (30)$$

Our EM algorithm starts with some initial estimate of the ability classes

$$\bar{\mathbf{Y}} := (\bar{y}_{it})_{(i,t) \in I \times T}.$$

To obtain $\bar{\mathbf{Y}}$, one may use the number of test items that each examinee answered correctly. The EM algorithm then repeats the E-step (expectation step) and M-step (maximization step) to maximize the log-likelihood function (30).

The M-step substitutes $\bar{\mathbf{Y}}$ into the log-likelihood function (30) and then maximizes it. We firstly consider maximizing Eq. (30a) subject to constraints (28). The method of Lagrange multipliers yields $\bar{\pi} := (\bar{\pi}_t)_{t \in T}$ defined by

$$\bar{\pi}_t \leftarrow \frac{\sum_{i \in I} \bar{y}_{it}}{|I|} \quad (31)$$

for each $t \in T$.

Next, we focus on maximizing Eq. (30b) after substituting Eq. (16) into it. This is equivalent to solving the smoothness-constrained model (21)–(27) with $\mathbf{Y} = \bar{\mathbf{Y}}$; that is, for each $j \in J$ we solve

$$\underset{\mathbf{s}_{j\cdot}, \mathbf{v}_{j\cdot}, \mathbf{w}_{j\cdot}}{\text{minimize}} \quad \sum_{(i,t) \in I \times T} \bar{y}_{it} \log(1 + \exp((1 - 2u_{ij})w_{jt})) \quad (32)$$

$$\text{subject to} \quad w_{jt} \leq w_{j,t+1} \quad (t \in T), \quad (33)$$

$$\sum_{t \in T} (s_{jt} + v_{jt}) \leq \gamma, \quad (34)$$

$$s_{jt} - v_{jt} = w_{j,t+2} - 2w_{j,t+1} + w_{jt} \quad (t \in T), \quad (35)$$

$$s_{jt} \geq 0, v_{jt} \geq 0 \quad (t \in T). \quad (36)$$

This problem minimizes a convex function subject to linear constraints, so it can be solved exactly and efficiently by standard nonlinear optimization software. The estimates obtained at this step are denoted by

$$\bar{\mathbf{W}} := (\bar{w}_{jt})_{(j,t) \in J \times T}.$$

The E-step updates $\bar{\mathbf{Y}}$ with its expected value based on the current estimates $\bar{\mathbf{W}}$ and $\bar{\pi}$. This amounts to assigning the posterior probability (29) as follows:

$$\bar{y}_{it} \leftarrow \frac{\bar{\pi}_t f(\mathbf{u}_{i\cdot} \mid \bar{\mathbf{w}}_{\cdot t})}{\sum_{\tau \in T} \bar{\pi}_\tau f(\mathbf{u}_{i\cdot} \mid \bar{\mathbf{w}}_{\cdot \tau})} \quad (37)$$

for all $(i, t) \in I \times T$. The E-step and M-step are repeated until a termination condition is satisfied. Our EM algorithm for solving the smoothness-constrained model (21)–(27) is summarized as follows:

- Step 0 (Initialization) Set $\bar{\mathbf{Y}}$ as an initial estimate, and go to Step 2.
 Step 1 (E-Step) Update $\bar{\mathbf{Y}}$ according to Eq. (37) for $(i, t) \in I \times T$.
 Step 2 (M-Step) Update $\bar{\pi}$ according to Eq. (31) for $t \in T$. Update $\bar{\mathbf{W}}$ by solving problem (32)–(36) for $j \in J$.
 Step 3 (Termination Condition) Terminate the algorithm if a termination condition is satisfied. Otherwise, return to Step 1.

4 Computational experiments

The computational results reported in this section evaluate the effectiveness of our smoothness-constrained NIRT model.

4.1 Experimental design

We evaluated the estimation accuracy of IRT models through the simulation process illustrated in Fig. 2.

In Step 1, we randomly generate θ_i for $i \in I$ from a standard normal distribution. Next, we give a true ability class t_i^* for $i \in I$ on the basis of θ_i and Table 1. For instance, if $-1.29 \leq \theta_i < -0.81$, we set $t_i^* := 2$. The ranges of θ in Table 1 were determined so that each ability class is assigned to approximately the same number of examinees.

Table 1 Relationship between ability class $t \in T$ and normal random number θ

t	range of θ	median of θ
1	$(-\infty, -1.29)$	-1.73
2	$[-1.29, -0.81)$	-1.02
3	$[-0.81, -0.49)$	-0.64
4	$[-0.49, -0.23)$	-0.36
5	$[-0.23, 0)$	-0.12
6	$[0, 0.23)$	0.12
7	$[0.23, 0.49)$	0.36
8	$[0.49, 0.81)$	0.64
9	$[0.81, 1.29)$	1.02
10	$[1.29, \infty)$	1.73

We used two types of function to create the true ICCs of test items. One was the two-parameter logistic (2PL) model

$$p_j^{2\text{PL}}(\theta) := \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))}, \quad (38)$$

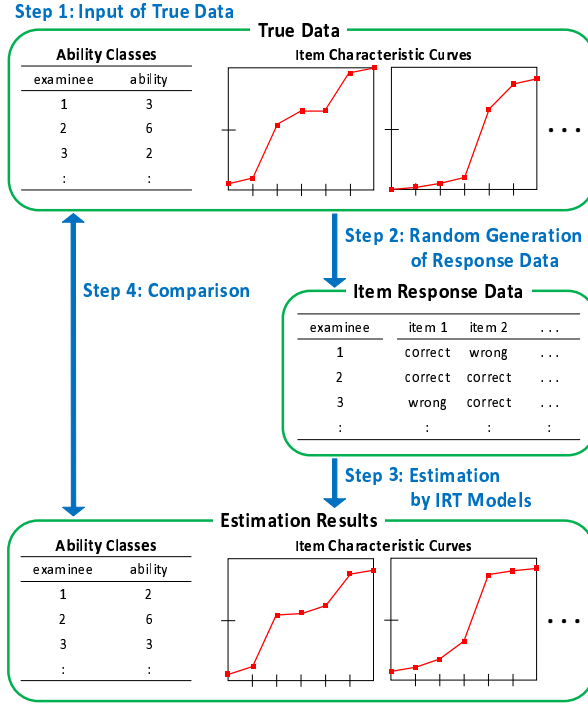


Fig. 2 Simulation process of model evaluation

where a_j and b_j are parameters of discrimination and difficulty, respectively. These parameters were drawn from uniform distributions for which $a_j \in [0.5, 2.0]$ and $b_j \in [-1.5, 1.5]$. Similarly to Nozawa (2008), the other type of function was the extended three-parameter normal ogive (3PN) model of order two:

$$p_j^{3PN}(\theta) := \Phi(a_{j2}(\theta - b_j)^3 + \sqrt{3a_{j1}a_{j2}}(\theta - b_j)^2 + a_{j1}(\theta - b_j)), \quad (39)$$

where Φ is the normal ogive, a_{j1} and a_{j2} are shape parameters, and b_j is a difficulty parameter. These parameters were drawn from uniform distributions for which $a_{j1} \in [0.4, 0.8]$, $a_{j2} \in [0.1, 0.5]$, and $b_j \in [-0.5, 0.5]$. The 3PN model is based on the assumption that examinees' abilities follow a bimodal distribution. Accordingly, the standard two-parameter logistic IRT models have difficulty in fitting ICCs defined by the 3PN model, whereas they can accurately fit ICCs defined by the 2PL model.

When the true ICC of the j th item was defined by the 2PL model (38), it was set as $x_{j1}^* := p_j^{2PL}(-1.73)$, $x_{j2}^* := p_j^{2PL}(-1.02)$, \dots , $x_{j,10}^* := p_j^{2PL}(1.73)$ according to the median of θ in each range (see Table 1). The true ICCs defined by the 3PN model (39) were set in the same manner. The percentage of ICCs defined by the 3PN model is denoted by ρ , where $\rho \in \{0\%, 20\%, 50\%\}$

similarly to Nozawa (2008). For instance, if $|J| = 60$ and $\rho = 20\%$, true ICCs of 12 items are created by the 3PN model.

In Step 2, binary item-response data \mathbf{U} are generated randomly based on the true ability classes and ICCs specified at Step 1. Specifically, examinees in the t th ability class answered the j th item correctly with probability x_{jt}^* .

In Step 3, ability classes and ICCs are estimated from the item-response data \mathbf{U} by using the following IRT models:

2PLM: two-parameter logistic model,

MHM: monotone-homogeneity model (6)–(10),

SCM(γ): smoothness-constrained model (21)–(27) with $\gamma \in \{0, 1, 2, 4\}$.

We used the `irt` package in R 3.1.2 (<http://www.R-project.org>) to perform the 2PLM computations. Here, the continuous-valued ability θ estimated by 2PLM was converted into an ability class t according to Table 1 for comparison purposes. To solve MHM and SCM, we implemented the EM algorithm by using MATLAB R2015a (<https://www.mathworks.com/products/matlab.html>), in which problem (32)–(36) was solved by the `fmincon` function in the MATLAB optimization toolbox. The initial estimate $\bar{\mathbf{Y}}$ was set by dividing examinees equally into 10 ability classes according to the number of correct answers. Every time $\bar{\mathbf{Y}}$ was updated in the EM algorithm, the ability classes of examinees were determined temporarily as $\hat{t}_i := \arg \max\{\bar{y}_{it} \mid t \in T\}$ for $i \in I$. The algorithm was terminated if \hat{t}_i remained the same as the previous one for all $i \in I$.

In Step 4, we evaluate the estimation accuracy of the IRT models by comparing the true data (Step 1) with the estimates (Step 3). Specifically, the root-mean-square error (RMSE) of the ability classes is calculated as

$$\sqrt{\frac{\sum_{i \in I} (t_i^* - \hat{t}_i)^2}{|I|}},$$

where \hat{t}_i is the estimated ability class. The RMSE of the ICCs is calculated as

$$\sqrt{\frac{\sum_{(j,t) \in J \times T} (x_{jt}^* - \hat{x}_{jt})^2}{|J||T|}},$$

where \hat{x}_{jt} is the estimated probability of a correct answer. We repeated these steps 10 times and show the average RMSEs in the following section.

4.2 Computational results

Tables 2 and 3 give the RMSEs of the ability classes and ICCs for the 12 experimental conditions. Here, the number of examinees was $|I| \in \{1000, 3000\}$, and the number of test items was $|J| \in \{30, 60\}$. Because the ordinal scale of neural test theory grades examinees into roughly 10 classes (Shojima 2007, 2008), the number of ability classes was $|T| = 10$. Note that the minimum RMSEs for each experimental condition are given in bold face in the tables.

Table 2 Root-mean-square error of item-characteristic curve estimation

$ I $	$ J $	ρ	2PLM	MHM	SCM(0)	SCM(1)	SCM(2)	SCM(4)
1000	30	0%	0.031	0.057	0.054	0.036	0.038	0.046
		20%	0.049	0.057	0.073	0.042	0.040	0.046
		50%	0.075	0.058	0.102	0.052	0.044	0.048
	60	0%	0.030	0.049	0.066	0.037	0.034	0.041
		20%	0.053	0.047	0.078	0.043	0.036	0.040
		50%	0.082	0.050	0.115	0.059	0.044	0.044
	3000	0%	0.017	0.041	0.056	0.025	0.026	0.034
		20%	0.046	0.041	0.069	0.032	0.028	0.034
		50%	0.073	0.041	0.102	0.044	0.033	0.034
	60	0%	0.031	0.034	0.064	0.027	0.023	0.028
		20%	0.056	0.034	0.080	0.038	0.027	0.030
		50%	0.083	0.034	0.116	0.058	0.033	0.029

We firstly focus on the accuracy of ICC estimation (Table 2). It is reasonable that 2PLM estimated ICCs very accurately when the percentage of 3PN ICCs was $\rho = 0\%$. However, the estimation accuracy of 2PLM was greatly reduced by increasing the percentage of 3PN ICCs. Indeed, when $\rho \geq 20\%$, the RMSEs were often smaller for MHM than for 2PLM. We also note that the estimation accuracy of MHM increased with the amount of item-response data. For instance, the RMSEs of MHM for $(|I|, |J|) = (1000, 30)$ were at least 0.057, and those for $(|I|, |J|) = (3000, 60)$ were 0.034.

The largest RMSEs of the ICCs were those for SCM(0) in almost all cases because the shapes of the ICCs were tightly restricted by the smoothness constraints with $\gamma = 0$. In contrast, SCM(1) frequently obtained higher estimation accuracy than did MHM, especially for $\rho \leq 20\%$. The smallest RMSEs were attained by SCM(2) in many cases, whereas only for $(|J|, \rho) = (60, 50\%)$ were those provided by SCM(4).

We next move on to the accuracy of ability-class estimation (Table 3). As in Table 2, the estimation accuracy of 2PLM reduced markedly with the percentage of 3PN ICCs. In contrast, MHM made relatively accurate estimates when $(|J|, \rho) = (60, 50\%)$. Very large RMSEs were still provided by SCM(0) as in Table 2. Meanwhile, the smallest RMSEs were often obtained by SCM(1) with $|J| = 30$, and those were obtained by SCM(2) with $|J| = 60$.

The results in Tables 2 and 3 confirm that smoothness constraints are very effective in improving the estimation accuracy of NIRT models. In particular, we may say that SCM(2) delivers the best estimation performance on the whole.

Table 4 gives the computation times (in seconds) required for estimating the IRT models. The computations of 2PLM were very rapid mainly because each ICC involved only two parameters. The computations of SCM(γ) became slower as γ or ρ increased. The computation time for estimating MHM was always the longest among all the models. These results suggest that when

Table 3 Root-mean-square error of ability-class estimation

$ I $	$ J $	ρ	2PLM	MHM	SCM(0)	SCM(1)	SCM(2)	SCM(4)
1000	30	0%	0.814	0.884	0.814	0.785	0.801	0.845
		20%	0.838	0.886	0.890	0.799	0.816	0.849
		50%	0.956	0.959	1.137	0.883	0.890	0.922
	60	0%	0.629	0.680	0.747	0.611	0.606	0.642
		20%	0.704	0.680	0.812	0.642	0.624	0.648
		50%	0.914	0.715	1.176	0.740	0.676	0.700
	3000	0%	0.808	0.877	0.841	0.787	0.803	0.850
		20%	0.831	0.872	0.881	0.792	0.801	0.841
		50%	0.982	0.929	1.150	0.871	0.869	0.905
	60	0%	0.642	0.629	0.739	0.585	0.576	0.603
		20%	0.747	0.642	0.828	0.638	0.599	0.621
		50%	0.997	0.694	1.190	0.761	0.666	0.675

Table 4 Computation times (s)

$ I $	$ J $	ρ	2PLM	MHM	SCM(0)	SCM(1)	SCM(2)	SCM(4)
1000	30	0%	3.3	172.2	42.6	43.3	75.6	140.2
		20%	3.2	173.0	56.7	62.5	83.7	129.4
		50%	3.1	200.8	77.0	106.3	115.6	158.2
	60	0%	8.2	383.1	81.8	113.2	118.6	186.5
		20%	8.2	317.4	144.0	166.7	149.0	214.9
		50%	8.4	409.6	224.6	265.2	220.0	259.9
	3000	0%	13.8	1140.0	88.3	128.3	310.2	597.9
		20%	14.2	1026.5	171.4	269.5	326.5	542.3
		50%	13.7	1273.7	181.6	281.2	323.0	672.6
	60	0%	36.7	1743.7	214.0	355.2	364.8	830.3
		20%	36.0	1949.8	443.8	765.2	555.2	825.2
		50%	35.8	3053.2	487.5	1008.0	898.0	1034.0

many 3PN ICCs are estimated subject to loose smoothness constraints, our EM algorithm takes a relatively long time to terminate.

Figs. 3 and 4 show illustrative examples of estimated ICCs together with the true ICCs. We firstly focus on Fig. 3, where $(|I|, |J|, \rho) = (3000, 60, 0\%)$, and the true ICCs were defined by the 2PL model. As expected, 2PLM fitted the true ICCs well. The MHM also fitted the true ICCs well, but it is noteworthy that the ICCs estimated by MHM moved around the true ICCs: for example, for the ability classes $t = 7, 8$, and 9 in Fig. 3(b) and $t = 3, 4$, and 6 in Fig. 3(c). The SCM(0) deviated partly from the true ICCs: for example, for the ability classes $t = 3, 4, 7$, and 8 in Fig. 3(a). In contrast, the true ICCs and those estimated by SCM(2) were almost the same because the nonparametric ICCs were made moderately less flexible by the smoothness constraints with $\gamma = 2$.

We next move on to Fig. 4, where $(|I|, |J|, \rho) = (3000, 60, 50\%)$ and the true ICCs were defined by the 3PN model. It is clear that neither 2PLM nor SCM(0) fitted the true ICC because these models create only logistic curves. In contrast, MHM and SCM(2) estimated the shapes of the true ICCs accurately, except that SCM(2) underestimated the probability of a correct answer for the ability class $t = 2$ in Fig. 4(a).

5 Conclusions

We devised an NIRT model with monotone homogeneity and smoothness constraints based on ordered latent classes. Our smoothness constraints avoid overfitting of nonparametric ICCs by smoothing them based on logistic curves. We also developed an EM algorithm for our smoothness-constrained NIRT model to estimate the nonparametric ICCs and the latent abilities of examinees efficiently.

The computational results demonstrated the effectiveness of our model in comparison to the two-parameter logistic model and the monotone-homogeneity model. Indeed, our model obtained more accurate estimation results than did the two-parameter logistic model when the latent abilities of examinees for some test items followed bimodal distributions. Moreover, our model outperformed the monotone-homogeneity model because of the effect of the smoothness constraints.

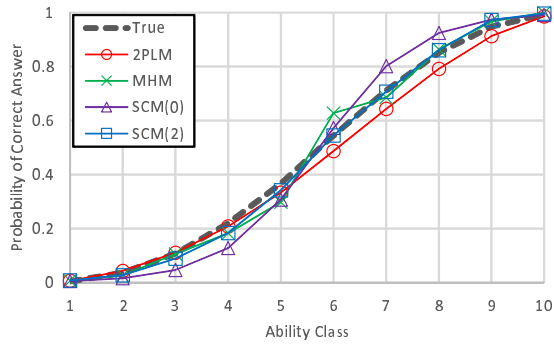
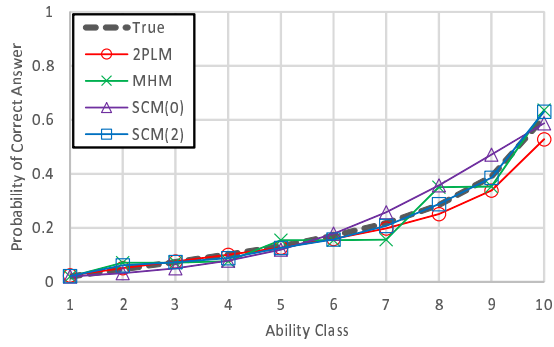
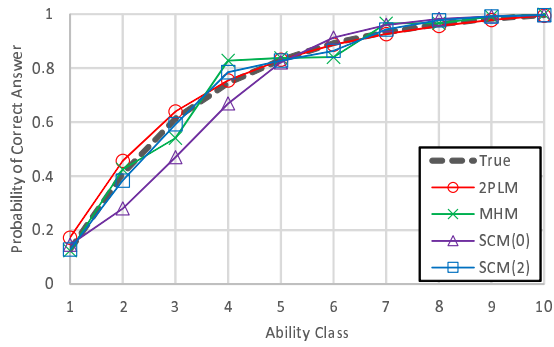
The contributions of this research are twofold. Firstly, we formulated the smoothness-constrained NIRT model as a mathematical optimization problem. Although we implemented the EM algorithm for the purpose of efficient computation, high-performance mixed-integer optimization algorithms could be used to compute an optimal solution to the problem (Bixby 2012). Secondly, we validated the utility of shape restrictions on nonparametric ICCs in avoiding an unstable ICC estimation specific to NIRT models. A future direction of study will be to impose other shape restrictions on nonparametric ICCs and evaluate their effectiveness.

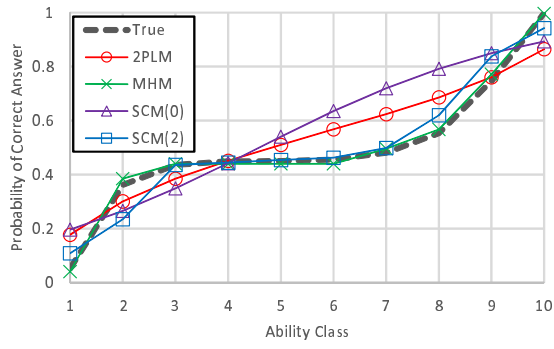
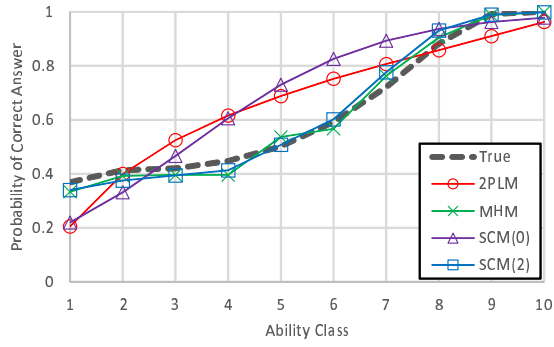
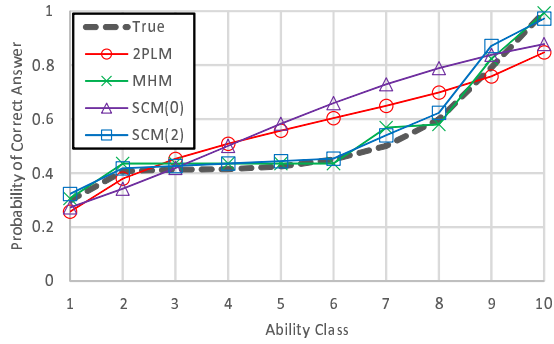
Although PIRT models are used commonly by many testing companies, it is known that they do not always fit the actual item-response data well. In contrast, NIRT models have the potential to estimate various forms of ICCs, and the estimation performance could be improved by incorporating the smoothness constraints. So, we expect that our research will extend the usefulness of NIRT models.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number 26750114.

References

- Bixby RE (2012) A brief history of linear and mixed-integer programming computation. In: Grötschel M (ed) *Documenta mathematica: Optimization stories*. Deutschen Mathematiker-Vereinigung, Berlin, pp 107–121

(a) test item $j = 7$ (b) test item $j = 9$ (c) test item $j = 27$ **Fig. 3** Estimated ICCs and true 2PL ICCs ($(|I|, |J|, \rho) = (3000, 60, 0\%)$)

(a) test item $j = 33$ (b) test item $j = 36$ (c) test item $j = 59$ **Fig. 4** Estimated ICCs and true 3PN ICCs ($(|I|, |J|, \rho) = (3000, 60, 50\%)$)

- Croon MA (1990) Latent class analysis with ordered latent classes. *Br J Math Stat Psychol* 43:171–192
- Croon MA (1991) Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *Br J Math Stat Psychol* 44:315–331
- Douglas J (1997) Joint consistency of nonparametric item characteristic curve and ability estimation. *Psychom* 62:7–28
- Douglas J, Cohen A (2001) Nonparametric item response function estimation for assessing parametric model fit. *Appl Psychol Meas* 25:234–243
- Falk CF, Cai L (2016) Semiparametric item response functions in the context of guessing. *J Educ Meas* 53:229–247
- Guo H, Sinharay S (2011) Nonparametric item response curve estimation with correction for measurement error. *J Educ Behav Stat* 36:755–778
- Johnson MS (2007) Modeling dichotomous item responses with free-knot splines. *Comput Stat Data Anal* 51:4178–4192
- Junker B, Sijtsma K (2001) Nonparametric item response theory in action: An overview of the special issue. *Appl Psychol Meas* 25:211–220
- Karabatsos G, Sheu CF (2004) Order-constrained Bayes inference for dichotomous models of unidimensional nonparametric IRT. *Appl Psychol Meas* 28:110–125
- Lee YS (2007) A comparison of methods for nonparametric estimation of item characteristic curves for binary items. *Appl Psychol Meas* 31:121–134
- Levy R (2009) The rise of Markov chain Monte Carlo estimation for psychometric modeling. *J Probab Stat* 53:7139
- Liang L, Browne MW (2015) A quasi-parametric method for fitting flexible item response functions. *J Educ Behav Stat* 40:5–34
- Liang T, Wells CS (2015) A nonparametric approach for assessing goodness-of-fit of IRT models in a mixed format test. *Appl Meas Educ* 28:115–129
- Liang T, Wells CS, Hambleton RK (2014) An assessment of the nonparametric approach for evaluating the fit of item response models. *J Educ Meas* 51:1–17
- Ligtvoet R, Vermunt JK (2012) Latent class models for testing monotonicity and invariant item ordering for polytomous items. *Br J Math Stat Psychol* 65:237–250
- Luzardo M, Rodríguez P (2015) A nonparametric estimator of a monotone item characteristic curve. In: van der Ark L, Bolt D, Wang WC, Douglas J, Chow SM (eds) *Quantitative psychology research*. Springer, Cham, pp 99–108
- Meijer RR, Tendeiro JN, Wanders RB (2015) The use of nonparametric item response theory to explore data quality. In: Reise SP, Revicki D (eds) *Handbook of item response theory modeling: Applications to typical performance assessment*. Routledge, pp 85–110
- Meredith W (1965) Some results based on a general stochastic model for mental tests. *Psychom* 30:419–440
- Mokken RJ (1971) *A theory and procedure of scale analysis with applications in political research*. New York, Berlin: Walter de Gruyter, Mouton

- Mokken RJ, Lewis C (1982) A nonparametric approach to the analysis of dichotomous item responses. *Appl Psychol Meas* 6:417–430
- Molenaar IW (2001) Thirty years of nonparametric item response theory. *Appl Psychol Meas* 25:295–299
- Monroe S, Cai L (2014) Estimation of a Ramsay-curve item response theory model by the Metropolis–Hastings Robbins–Monro algorithm. *Educ Psychol Meas* 74:343–369
- Mori J, Kano Y (2013) Item response theory using a finite mixture of logistic models with item-specific mixing weights. *J Jpn Soc Comput Stat* 26:17–38
- Nozawa Y (2008) Comparison of parametric and nonparametric IRT equating methods under the common-item nonequivalent groups design. Dissertation, University of Iowa
- Ramsay JO (1988) Monotone regression splines in action. *Stat Sci* 3:425–441
- Ramsay JO (1991) Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychom* 56:611–630
- Ramsay JO, Abrahamowicz M (1989) Binomial regression with monotone splines: A psychometric application. *J Am Stat Assoc* 84:906–915
- Ramsay JO, Silverman BW (2005) *Functional data analysis*. Springer, New York
- Ramsay JO, Winsberg S (1991) Maximum marginal likelihood estimation for semiparametric item analysis. *Psychom* 56:365–379
- Rossi N, Wang X, Ramsay JO (2002) Nonparametric item response function estimates with the EM algorithm. *J Educ Behav Stat*, 27:291–317
- Shojima K (2007) Neural test theory. DNC Research Note 07-02, The National Center for University Entrance Examinations
- Shojima K (2008) Neural test theory: A latent rank theory for analyzing test data. DNC Research Note 08-01, The National Center for University Entrance Examinations
- Sijtsma K, Molenaar IW (2002) *Introduction to nonparametric item response theory*. Sage, Thousand Oaks
- Simonoff JS (2012) *Smoothing methods in statistics*. Springer, New York
- Stout WF (1987) A nonparametric approach for assessing latent trait unidimensionality. *Psychom* 52:589–617
- Takano Y, Tsunoda S, Muraki M (2016) Mathematical optimization models for nonparametric item response theory. *Inf Sci Appl Math* 23:1–18
- van der Linden WJ, Hambleton RK (eds) (2013) *Handbook of modern item response theory*. Springer, New York
- van Onna MJH (2002) Bayesian estimation and model selection in ordered latent class models for polytomous items. *Psychom* 67:519–538
- Vermunt JK (2001) The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Appl Psychol Meas* 25:283–294